Blind Source Separation of Sources with Different Magnitudes

J.-P. Nadal[1], E. Korutcheva[2] and F. Aires[3]
[1] *Laboratoire de Physique Statistique de l'ENS*,*
Ecole Normale Supérieure,
24, rue Lhomond - 75231 Paris Cedex 05, France
[2] *Departamento de Física Fundamental,*
Universidad Nacional de Educación a Distancia (UNED),
c/Senda del Rey s/n - 28080 Madrid, Spain†
[3] *NASA/ Goddard Institute for Space Studies,*
2880 Broadway, New York, NY 10025, USA

*Laboratoire associé au CNRS, à l'ENS et aux Universités Paris 6 et Paris 7.
†Permanent address: G.Nadjakov Inst. Solid State Physics, Bulgarian Academy of Sciences, 1784 Sofia, Bulgaria

Abstract- **We investigate the information processing of a linear mixture of independent sources of different magnitudes when a number m of the sources can be considered as "strong" as compared to the other ones, the "weak" sources. We find that it is preferable to perform blind source separation in the space spanned by the strong sources, first projecting the signal onto the m largest principal components. We illustrate the analytical results with numerical simulations on real data.**

I. INTRODUCTION

During the recent years many studies have been devoted to the study of Blind Source Separation (BSS) and more generally of Independent Component Analysis (ICA) (see e.g.[7,5,6,2]). Within the standard framework one assumes a multidimensional measured signal to result from a linear mixture of statistically independent components, or "sources". In most cases one makes the optimistic hypotheses that the number of sources is equal to the dimension of the signal (the number of captors), and that the unknown mixture matrix is invertible. The goal of BSS is then to compute an estimate of the inverse of the mixture matrix in order to extract from the signal the independent components.

In the present paper we study the effect of having sources with different "strengths" when performing BSS. After giving a proper definition of the strength of a source, the main purpose of our study is to relate the strength of a source to its contribution to the information conveyed by the processing system about the signal, and to consider with more details the case where some of the sources are very weak compared to the others. We will show that in that case it is worthwhile to project the data onto the space generated by the strong sources in order to extract meaningful information and to avoid numerical problems. The contributions to the (projected) signal from the weak sources can then be considered as noise terms added to the linear mixture of strong sources. Since the sources are independent, this "noise" is thus independent of the "pure" signal (the part due to the strong sources).

II. THE MODEL

We consider the information processing of a signal which is a N-dimensionnal linear mixture of N independent sources. At each time t one observes $\mathbf{S}(t) = \{S_j(t), j = 1, ..., N\}$ which can be written in term of the unknown sources $\mathbf{s}(t) = \{s_\alpha(t), \alpha = 1, ..., N\}$ as:

$$S_j = \sum_{\alpha=1}^{N} M_{j\alpha}\, s_\alpha, \qquad j = 1, ..., N. \tag{1}$$

where $\mathbf{M} = \{M_{j\alpha}, j = 1, ..., N, \alpha = 1, ..., N\}$ is the mixture matrix assumed to be invertible. As it is well known, and easily seen from the above equation, it is not possible to distinguish between the mixture of $\mathbf{s}$ with the matrix $\mathbf{M}$ from the mixture of $\mathbf{s}' \equiv \mathbf{PDs}$ with the matrix $\mathbf{M}' \equiv \mathbf{MD}^{-1}\mathbf{P}^{-1}$ where $\mathbf{D}$ is an arbitrary diagonal matrix with non zero diagonal elements, and $\mathbf{P}$ an arbitrary permutation of N indices. If we decide to consider both normalized sources and normalized mixture matrices, we are left with a diagonal matrix $\mathbf{D}$ which defines the "strengths" of the sources. More precisely we write

$$S_j = \sum_{\alpha=1}^{N} \overline{M}_{j\alpha}\, \eta_\alpha\, s_\alpha, \qquad j = 1, ..., N. \tag{2}$$

assuming zero mean and unit variance for every source:

$$< s_\alpha >= 0, \qquad < s_\alpha^2 >= 1, \qquad \alpha = 1, ..., N \tag{3}$$

where $< . >$ denotes the average with respect to the (unknown) sources probability distributions,

$$\rho(\mathbf{s}) \;=\; \prod_\alpha \rho_\alpha(s_\alpha), \tag{4}$$

and with $\overline{\mathbf{M}}$ the normalized mixture matrix. The normalization can be chosen in different ways, and two of them are of particular interest for what follows. The simplest one is, for each α,

$$\left[\overline{\mathbf{M}}^{\mathbf{T}}\overline{\mathbf{M}}\right]_{\alpha\alpha} \;=\; \sum_{j=1}^{N} (\overline{M}_{j\alpha})^2 = 1. \tag{5}$$

The second one is a normalization on the inverse of the mixture matrix:

$$\left[\overline{\mathbf{M}}^{-\mathbf{1}}\overline{\mathbf{M}}^{\mathbf{T-1}}\right]_{\alpha\alpha} \;=\; \sum_{j=1}^{N} (\left[\overline{M}^{\,-1}\right]_{\alpha j})^2 = 1. \tag{6}$$

Once a particular normalization, such as (5) or (6), is chosen, the parameters η_α in (2) are well defined and can be understood as the relative strengths of the sources.

The purpose of this investigation is to show that it is worthwhile to project the data on the space generated by the strong sources in order to extract meaningful information and to avoid numerical problems. The contributions to the signal from the weak sources can then be considered as noise terms added to the linear mixture of strong sources.

III. INFORMATION CONTENT OF THE DATA

Let us now compute the amount of information conveyed by the data, $\mathbf{S}$, about the sources, that is the mutual information[4] $I(\mathbf{S}, \mathbf{s})$. To do so we consider

$$S_j = \sum_{\alpha=1}^{N} \overline{M}_{j\alpha}\eta_\alpha s_\alpha \; + \; \nu_j, \qquad j = 1, ..., N. \quad (7)$$

where $\boldsymbol{\nu} = \{\nu_j, j = 1, ..., N\}$ is a vanishing additive noise, $< \nu_j >= 0, < \nu_j \; \nu_k >= b \; \delta_{j,k}$ with $b \to 0$. Then $I(\mathbf{S}, \mathbf{s})$ is a constant (that is a quantity that depends on b alone) plus the data entropy. Since the mixture matrix is invertible, we have

$$I(\mathbf{S}, \mathbf{s}) = cst. + \ln|\det \overline{\mathbf{M}}| + \sum_\alpha \ln \eta_\alpha - \sum_\alpha \int dh_\alpha \rho_\alpha(h_\alpha) \ln \rho_\alpha(h_\alpha). \quad (8)$$

The last term in the above expression is the source entropies. One should remember that the s's are the normalized sources, $< s_\alpha^2 >= 1$. This shows that each source contributes to the information by a combination of its strength and its entropy: the strength term favors strong sources, whereas the entropy term favors the sources with a p.d.f. close to Gaussian. The entropy of a source cannot exceeds the one of a Gaussian with same variance, that is

$$- \int dh_\alpha \rho_\alpha(h_\alpha) \ln \rho_\alpha(h_\alpha) \leq \frac{1}{2} \ln 2\pi e. \quad (9)$$

Hence the information can be easily dominated by the strength term, which is not bounded.

It is known that for performing BSS perfect knowledge of the sources distribution is not necessary, and working on the cumulants of order 2 and 3 or 4 is sufficient (see e.g.[6,10]). We can thus analyze the result (8) by making a close-to-Gaussian approximation[6,10]. If we assume the sources to have non zero third order cumulants,

$$\lambda_\alpha^{(3)} \equiv < s_\alpha^3 >_c, \quad (10)$$

we replace the source distribution ρ_α by

$$\hat{\rho}_\alpha(s_\alpha) = \frac{e^{-s_\alpha^2/2}}{\sqrt{2\pi}} \left(1 + \lambda_\alpha^{(3)} \frac{s_\alpha(s_\alpha^2 - 3)}{6}\right). \quad (11)$$

The distribution $\hat{\rho}_\alpha$ has the same three first moments as the true distribution ρ_α[1]. Within this approximation the mutual information (8) reads

$$I(\mathbf{S}, \mathbf{s}) = cst. \; + \; \ln|\det \overline{\mathbf{M}}| \; + \; \sum_\alpha \ln \eta_\alpha + \frac{\mathbf{N}}{\mathbf{2}} \ln \mathbf{2\pi e} - \frac{1}{12} \sum_\alpha \; < s_\alpha^3 >_c^2 . \quad (12)$$

From the above expression the most important source are those for which the quantity

$$< s_\alpha^3 >_c^2 \; - \; \ln \eta_\alpha \quad (13)$$

is the smallest.

IV. CHARACTERIZATION FROM THE INFOMAX PRINCIPLE

The Infomax criterion[8,9] will allow us to get some more insight onto the link between the sources strengths and the amount of information that can be extracted from the data.

We consider the information processing of the signal by a nonlinear network, and we are interested in computing the mutual information $I(\mathbf{V}, \mathbf{S})$ between the input $\mathbf{S}$ and the output $\mathbf{V} = \{V_i, i = 1, ..., N\}$ of the network. Since the signal is a linear mixture the relevant architecture is a linear processing followed by a (possibly) nonlinear transfer function which may differ from neuron to neuron:

$$V_i = f_i(h_i) \; + \; \nu_i \quad (14)$$

$$h_i = \sum_j \; J_{ij} \; (\; S_j \; + \; \nu_j^0 \;), \quad (15)$$

where $\boldsymbol{\nu}_0 = \{\nu_j^0, j = 1, ..., N\}$ and $\boldsymbol{\nu} = \{\nu_i, i = 1, ..., N\}$ are additive input and output noise, respectively, with $< \boldsymbol{\nu}_0 >= 0, < \boldsymbol{\nu} >= 0, < \nu_j^0 \; \nu_{j'}^0 >= b^0 \; \delta_{j,j'}, < \nu_i \; \nu_{i'} >= b \; \delta_{i,i'}$. The J_{ij} can be viewed as synaptic efficacies and the h_i's as post-synaptic potentials (PSP). As explained in the previous section, the noise has to be introduced in order to have a finite mutual information, and we take the limit $0 \leq b^0 << b << 1$. For strictly zero input noise, $b^0 = 0$, in the limit $b \to 0$ the mutual information is up to a constant equal to the output entropy. As shown in[9] its maximization over the choice of both $\mathbf{J}$ and the transfer functions f_i's leads to BSS. One can then derive practical algorithms for performing BSS[3]. In this limit of $b^0 = 0$ all the sources play the same role, that is the maximum of the mutual information is independent of the individual sources properties as well as of the mixture matrix. When one takes into account a non zero input noise, then at first non trivial order in $\frac{b^0}{b}$ one sees that the input noise introduces a scale witch breaks this invariance. More precisely, at first order in $\frac{b^0}{b}$ the mutual information can be written (see[9] for details):

$$I(\mathbf{V}, \mathbf{S}) \; = \; I_0(\mathbf{V}, \mathbf{S}) \; - \; \frac{b^0}{2b} \sum_{i=1}^{N} C_{ii} \int dh_i \psi_i(h_i) f_i^{'2} \quad (16)$$

where $I_0(\mathbf{V}, \mathbf{S})$ is the value at $b^0 = 0$,

$$I_0(\mathbf{V}, \mathbf{S}) \; = \; cst. \; - \int d\mathbf{h} \psi(\mathbf{h}) \ln \frac{\psi(\mathbf{h})}{\prod_{i=1}^N f_i'(h_i)} \quad (17)$$

and $\frac{b^0}{b} \; C_{ii}$ is the variance of the noise on the PSP h_i:

$$C_{ii} \equiv \left[\mathbf{J}\mathbf{J}^T\right]_{ii} \quad (18)$$

Finally, $\psi(\mathbf{h})$ is the probability distribution of $\mathbf{h}$ induced by the sources input distribution, and $\psi_i(h_i)$ the marginal distribution of the PSP h_i. At a given $\mathbf{J}$, optimizing with respect to the choice of transfer functions gives

$$f_i'(h_i) = \psi_i(h_i) \; \{ \; 1 \; + \; \frac{b^0}{b} \, C_{ii} \; \left[< \psi_i^2 > - \psi_i^2(h_i) \right] \; \} \quad (19)$$

with $< \psi_i^2 > = \int dh_i \psi_i(h_i) \psi_i^2(h_i) = \int dh_i \psi_i(h_i)^3$.

We now optimize over $\mathbf{J}$. At zeroth order the optimum is reached for $\mathbf{J} = \mathbf{M}^{-1}$ (up to an arbitrary permutation), so that we write

$$\mathbf{W} \equiv \mathbf{J}\mathbf{M} = \mathbf{1}_N + \frac{b^0}{b} \mathbf{W}^1 \tag{20}$$

where $\mathbf{1}_N$ is the $N \times N$ identity matrix. Expending the mutual information at first order in $\frac{b^0}{b}$ one finds that there is no contribution from $\mathbf{W}^1$ to this order. Hence the mutual information at first order in $\frac{b^0}{b}$ is given by (16) at $\mathbf{J} = \mathbf{M}^{-1}$, with f_i' given by (19) in which we set $\psi_i = \rho_i$. This gives

$$I(\mathbf{V}, \mathbf{S}) \; = \; cst. \; - \; \frac{b^0}{2b} \sum_{\alpha=1}^{N} C_{\alpha\alpha} \; \int ds_\alpha [\rho_\alpha(s_\alpha)]^3 \tag{21}$$

with

$$C_{\alpha\alpha} \; = \; \left[\mathbf{M}^{-1} \mathbf{M}^{T-1} \right]_{\alpha\alpha} \tag{22}$$

One sees that the term depending on $\mathbf{M}$ is what appears in the normalization (6) of the mixture matrix. Hence if we choose this particular normalization (6) in order to define the strengths η_α of the sources, one can rewrite

$$I(\mathbf{V}, \mathbf{S}) \; = \; cst. \; - \; \frac{b^0}{2b} \sum_{\alpha=1}^{N} \; \frac{1}{\eta_\alpha^2} \; < \rho_\alpha^2 > \tag{23}$$

with $< \rho_\alpha^2 > = \int ds_\alpha \rho_\alpha(s_\alpha)^3$. The above expression shows how each source α contributes to the mutual information in term of its strength η_α and its pdf ρ_α.

Within the close-to-Gaussian approximation (11) one gets

$$I(\mathbf{V}, \mathbf{S}) \; = \; cst \; - \; \frac{b^0}{b} \sum_{\alpha=1}^{N} \; < s_\alpha^3 >_c^2 \; \frac{1}{\eta_\alpha^2}. \tag{24}$$

Hence the sources which contribute the most to the conveyed information are those for which the quantity

$$\mathcal{E}_\alpha \equiv < s_\alpha^3 >_c^2 \; \frac{1}{\eta_\alpha^2} \tag{25}$$

is the smallest. One should remember that η_α is given by

$$\frac{1}{\eta_\alpha^2} = \sum_{j=1}^{N} \left([\mathbf{M}^{\,-1}]_{\alpha j} \right)^2. \tag{26}$$

V. BSS WITH NOISY DATA

A standard approach in data processing consists in projecting the data onto the eigenspace associated with the largest eigenvalues. In the present context of BSS, it is reasonable to expect the space spanned by the strong sources to be essentially the same as the one associated to the largest principal components. It is the purpose of this section to give a positive and more precise answer to this question.

We consider the specific case where m source are "strong", while $N - m$ sources are "weak". More precisely, defining ϵ as a small parameter, $\epsilon << 1$, we assume

$$\begin{aligned} &\eta_\alpha \sim O(1 = \epsilon^0) \text{ for } \alpha = 1, ..., m \\ &\eta_\alpha \sim O(\epsilon) \text{ for } \alpha = m+1, ..., N. \end{aligned} \tag{27}$$

Let us now assume that we have preprocessed the data by projecting it onto the m largest principal components. Instead of the model (1) we have thus to consider the model

$$S_j = \sum_{\alpha=1}^{m} M_{j\alpha} s_\alpha \; + \; \nu_j^0, \qquad j = 1, ..., m \tag{28}$$

where $\mathbf{M}$ is now a $m \times m$ invertible mixture matrix, such that $\mathbf{M}\mathbf{M}^T$ has m non zero, of order $1 = \epsilon^0$, eigenvalues. The s_α's ($\alpha = 1, ..., m$) are the sources of interest, and the ν_j^0's are additive noises, resulting from the weak sources, as explained in the previous section. This noise $\boldsymbol{\nu}_0 = \{\nu_j^0, j = 1, ..., m\}$ is uncorrelated with the m (strong) sources, and of arbitrary distribution $P(\boldsymbol{\nu}_0)$. Since we are working in the small ϵ regime, all we will need is to characterize this distribution by its first two cumulants:

$$\begin{aligned} &< \boldsymbol{\nu}_0 > = 0 \\ &< \boldsymbol{\nu}_0 \boldsymbol{\nu}_0^T > = \epsilon^2 \, \mathbf{B}, \end{aligned} \tag{29}$$

where $\mathbf{B}$ is a (possibly non diagonal) $m \times m$ symmetric matrix. The problem we are considering now is thus strictly the same as the one of performing BSS on a linear mixture of m sources corrupted by some additive input noise, which, although small, cannot be neglected.

Let us now consider this noisy BSS problem within the Infomax approach in the line of[9]. The network we consider has the same architecture as the one defined in (15), but with m inputs and outputs:

$$V_i = f_i(h_i) \; + \; \nu_i \tag{30}$$

$$h_i = \sum_{j=1}^{m} \; J_{ij} \; (\; S_j \; + \; \nu_j^0 \;) \quad i = 1, ..., m, \tag{31}$$

with $< \nu_i \; \nu_{i'} > = b \; \delta_{i,i'}$. The limit to be considered here is the one of a vanishing output noise, $b \to 0$, but at a given input noise level:

$$0 < b << \epsilon^2. \tag{32}$$

Another important difference with the calculation done in section IV, is that here we are interested in computing the information conveyed about the global input, $\mathbf{S}+\boldsymbol{\nu}_0$, and not about the "pure" signal alone $\mathbf{S}$. This is because *we* have decided to call "signal" the strong sources and "noise" the weak sources, whereas in section IV the input noise would correspond to some noise at the level of the receptors.

In this limit of vanishing output noise, the mutual information $I(\mathbf{V}, \mathbf{S}+\boldsymbol{\nu}_0)$ between the output and the input of the network is up to a constant equal to the output entropy. To simplify the analysis, we assume a full adaptation of the transfer functions, which means[9], for $\mathbf{J}$ given,

$$f_i'(h_i) = \psi_i(h_i), i = 1, ..., m, \tag{33}$$

where $\psi_i(h_i)$ is the marginal probability distribution of the PSP h_i. As a result, the mutual information is up to a constant equal to the redundancy between the PSP's[9]:

$$I(\mathbf{V}, \mathbf{S}) = Const - \int d^m\mathbf{h}\psi(\mathbf{h}) \ln \frac{\psi(\mathbf{h})}{\prod_{i=1}^m \psi_i(h_i)}. \tag{34}$$

In term of the sources distributions, the distribution $\psi(\mathbf{h})$ is given by:

$$\psi(\mathbf{h}) = \int \prod_{\alpha=1}^{m} ds_\alpha \, \rho_\alpha(s_\alpha) \int d^m\boldsymbol{\nu}_0 P(\boldsymbol{\nu}_0) \prod_i \delta(h_i - \sum_\alpha [JM]_{i\alpha} \, s_\alpha \; - \; \sum_j J_{ij} \, \nu_j^0) \tag{35}$$

Since in (35) the noises ν_j^0 are $\sim O(\epsilon)$ we can perform an expansion, leading to the following expression:

$$\psi(\mathbf{h}) = \{ \; 1 \; + \; \epsilon^2 \sum_{i,i'} \left[\mathbf{JBJ}^T\right]_{ii'} \, \partial_i \partial_{i'} \; \} \; \psi^0(\mathbf{h}), \tag{36}$$

where ∂_i means the partial derivative with respect to h_i, and $\psi^0(\mathbf{h})$ is the p.d.f. that would be obtained at $\epsilon = 0$.

We consider now the maximization of the mutual information over the choice of $\mathbf{J}$, taking into account that ϵ is small. If ϵ was strictly zero, we would be back to the noiseless BSS problem for which the optimum is reached for $\mathbf{J} = \mathbf{M}^{-1}$ (up to an arbitrary permutation). So for nonzero ϵ we write

$$\mathbf{W} \equiv \mathbf{JM} = \mathbf{1}_m + \mathbf{W}^1, \tag{37}$$

where $\mathbf{1}_m$ is the $m \times m$ identity matrix, and $\mathbf{W}^1$ a matrix of order at least ϵ. ψ^0 can then be written as

$$\psi^0(\mathbf{h}) = \left[\prod_\alpha \rho_\alpha(h_\alpha)\right] \; \{1 - \sum_\alpha [\ln \rho_\alpha]' \sum_\beta W^1_{\alpha\beta} h_\beta - Tr\mathbf{W}^1 + \epsilon^2 ...\} \tag{38}$$

and similarly, for the marginal distributions:

$$\psi^0_\alpha(h_\alpha) = \rho_\alpha(h_\alpha) \; \{1 - [\ln \rho_\alpha]' W^1_{\alpha\alpha} h_\alpha - W^1_{\alpha\alpha} + \epsilon^2 ...\} \tag{39}$$

The substitution of (35) in (34) gives for the mutual information:

$$I(\mathbf{V}, \mathbf{S}) = I_0(\mathbf{V}, \mathbf{S}) - \frac{\epsilon^2}{2} \{ \sum_{i,i'} B_{ii'} \int \mathbf{h} \frac{\partial^2 \psi^0}{\partial h_i \partial h_i'} \ln \psi^0 - \sum_i B_{ii} \int dh_i \frac{\partial^2 \psi_i^0}{\partial h_i^2} \ln \rho_i \}. \tag{40}$$

with $\psi^0(\mathbf{h}) = \prod_i \psi_i^0(h_i) = \prod_i \rho_i(h_i)$. The term $I_0(\mathbf{V}, \mathbf{S})$ corresponds to the part of the mutual information which does not take into account the "weak" sources. It is the same as if one computes the mutual information between the output $\mathbf{V}$ and the signal $\mathbf{Ms}$; $I(\mathbf{V}, \mathbf{Ms})$.

It is easy to be shown that the two contributions of order ϵ^2 cancel and that corrections due to the weak sources appear in order ϵ^4. The last permits to work, up to high approximation, using only the part corresponding to the "strong" sources.

We tested the above analysis on a toy example by considering the ICA of natural images performed in[3]. First we reproduced the results in[3]. We then created a new data base with artificially increased component strengths: new images are computed as a linear mixture of the previous ICA basis function but the strength of 20 components was augmented 100 times compared to the other 124. We performed ICA in this new data base, with the same algorithm based on Infomax[9,3], but after projecting the data onto the 20 largest principal components. We showed that the computational time within this scheme is considerably decreased.

VI. CONCLUDING REMARKS

We have discussed the task of Blind Source Separation in the case of a mixture of sources of unequal strengths.

We have presented different ways of defining the relative strengths of the sources. In particular, when non zero input noise is taken into account the contribution of a source to the conveyed information can be characterized by a criterion which combines the mixture matrix elements and the third cumulant of the source distribution. This allows to define the strength of a source once a proper normalization of the mixture matrix is assumed.

The analysis indicates also that, although arbitrary, the assumed normalization of the mixture matrix may have an important practical role in the analysis of the outcome of an ICA, whenever one wants to extract the "meaningful" sources. Which part of the signal is more important is of course an application dependent notion. Prior knowledge related to a given case should allow to define the proper normalization from which the appropriate scale of source strengths can be defined. Conversely each chosen normalization implies a particular physical interpretation which should be kept in mind when analyzing the outcome of an ICA.

We have considered with more details the particular case of the information processing of a linear mixture of independent sources when some of them are very weak as compared to the other sources. One should note that in such case the notion of strong *versus* weak is independent of the mixture matrix normalization. It is easily seen that the presence of weak sources leads to an almost singular mixture matrix, and this manifests itself by the existence of very small eigenvalues in the PCA analysis. We have shown that it is relevant to project the input data onto the largest principal components in order to extract the strongest independent sources.

We illustrated this result on the ICA of the image data base studied in[3].

ACKNOWLEDGMENTS

We thank T. Bell for giving us access to his image data base. This work has been partly supported by the French contract DGA 962557A/DSP. E.K. warmly thanks for hospitality the Laboratoire de Physique Statistique de l'Ecole Normale Supérieure, where this work was performed. E.K. has been also supported by the Spanish DGES project PB97-0076 and partly by the French-Spanish Program PICASSO and the Bulgarian Scientific Foundation Grant F-608.

---

[1] M. Abramowitz and I. A. Stegun 1972, *Handbook of Mathematical Functions* (Dover).

[2] S.-I. Amari and J.-F. Cardoso 1997, Blind source separation-semiparametric statistical approach, *IEEE Tr. on Signal Processing*, 45(11):2692, Special issue on neural networks.

[3] A. Bell and T. Sejnowski 1995, An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159.

[4] R. E. Blahut 1988, *Principles and Practice of Information Theory* (Addison-Wesley, Cambridge MA).

[5] J.-F. Cardoso 1989, Source separation using higher-order moments, in *Proc. Internat. Conf. Acoust. Speech Signal Process., Glasgow 1989*, pp. 2109–2112.

[6] P. Comon 1994, Independent component analysis, a new concept ? *Signal Processing*, 36:287–314.

[7] J. Herault, C. Jutten and B. Ans 1985, Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé, in *Proc. GRETSI, Nice 1985* pp. 1017-1020.

[8] R. Linsker 1988, Self-organization in a perceptual network. *Computer*, 21:105–17.

[9] J.-P. Nadal and N. Parga 1994, Nonlinear neurons in the low-noise limit: a factorial code maximizes information transfer *Network: Computation in Neural Systems*, 5:565–581.

[10] J.-P. Nadal and N. Parga 1997, Redundancy reduction and independent component analysis: Algebraic and adaptive approaches *Neural Computation*, 9(7):1421–1456.